12th International Conference on Information Fusion
Seattle, WA, USA, July 6-9, 2009

Model Based Fusion and Services Automation Support

Mark A. Thomas
Tactical Information Fusion Branch
U.S. Army Research Laboratory
Aberdeen Proving Ground, MD
mark.thomas3@us.army.mil

Abstract – *This paper will report on ongoing research and development to support automated fusion at the U.S. Army Research Laboratory centered on human-collected information and processing. The overall approach for our research is three pronged: exploit soft information sources such as human-generated reports and open source information, develop discrete services, and focus on the end-user.*

Soft information sources provide information on relationships between individuals, organizations, locations, time, and events. Extracting this information and structuring it for efficient computer processing is a challenge which must be overcome to provide better analytical support to the user.

Discrete services provide for the development of components which can be composed into a system. Processing services such as text extractors, web crawlers, and graph processors within a Services-Oriented Architecture provides modularity and scalability.

The focus on the end-user grounds the research in the real needs of soldiers on the ground in the near-term and the future.

Keywords: automated fusion, services, human sensors.

Introduction

Current counter-insurgency operations reinforce the need for efficient processing of human authored information **[1].** Information gathered from people is the basis of most intelligence in counter-insurgency operations. **[1].** Face-to-face interviews, interrogation reports, and overheard chatter are traditional sources of information used to identify, track, and develop associations between individuals, link them to events, and develop lists of key individuals in organizations, or High Value Individuals (HVI)**. [1]** The research described in this paper is an attempt to understand and utilize the rich content of human collected information for better situational awareness, the Joint Directors of Laboratories Level 2 in the Fusion Model **[2].**

The U.S. Army collects hundreds of human-generated reports per day, shared through semi structured messages. The content of these messages, by and large, are unstructured. Early U.S. Army digital devices, the AN/PSG 22A (DMD), used for Field Artillery Calls for Fire **[3],** created and transmitted highly structured messages. Each field has an enumerated type entered using a menu. But even the DMD provided the ubiquitous "Freetext" message to allow the user to transmit information of interest outside of the confines of the structured messages **[3].**

Given the availability and nature of human generated information, the issue of processing this data becomes important. What processing framework is required? How can one represent the data for computer reasoning? How can one present the data to the user in the most understandable fashion? ARL is researching these issues, and is developing a prototype to demonstrate and provide a means to assess the capabilities of newly developed computational tools as they become available.

Processing Framework

With an eye to the end-user's needs, let us postulate the following scenario. Given a large corpus of data collected by human agents, an analyst may seek the following information: what is the relationship between actor A and event B? To investigate this issue, the analyst would read multiple reports, create a link diagram, and document the evidence to support any conclusions and recommendations. Automation assistance in this workflow can occur at multiple points. First is the gathering of relevant information. A user interface with a query capability will suffice. The power and flexibility of Google® is an example of how easy it is to gather information with a few keywords or phrases. The data must be in a suitable form for such a search to work automatically, however, and this is the problem with unstructured information. Past research in extracting information from HUMINT messages demonstrates the difficulty in extracting information from unstructured data **[4][5].** The second form of automation assistance is the development of link diagrams. The semantic web technology and Resource Description Framework (RDF) Ontology Web Language (OWL) can assist this process **[6].** Lastly, the system can assist by visualization of the results to the user. The Services Oriented Architecture used in our project is the Distributed Common Ground Station-Army (DCGS-A) DCGS-A Application

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **JUL 2009**	2. REPORT TYPE	3. DATES COVERED **06-07-2009 to 09-07-2009**

4. TITLE AND SUBTITLE **Model Based Fusion and Services Automation Support**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Tactical Information Fusion Branch,U.S. Army Research laboratory, ,Aberdeen Proving Ground,MD**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

See also ADM002299. Presented at the International Conference on Information Fusion (12th) (Fusion 2009). Held in Seattle, Washington, on 6-9 July 2009. U.S. Government or Federal Rights License.

14. ABSTRACT

This paper will report on ongoing research and development to support automated fusion at the U.S. Army Research Laboratory centered on human-collected information and processing. The overall approach for our research is three pronged: exploit soft information sources such as human-generated reports and open source information, develop discrete services, and focus on the end-user. Soft information sources provide information on relationships between individuals, organizations locations, time, and events. Extracting this information and structuring it for efficient computer processing is a challenge which must be overcome to provide better analytical support to the user. Discrete services provide for the development of components which can be composed into a system. Processing services such as text extractors, web crawlers, and graph processors within a Services- Oriented Architecture provides modularity and scalability. The focus on the end-user grounds the research in the real needs of soldiers on the ground in the near-term and the future.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Public Release**	**4**	

Framework (DAF). The DAF provides an interface and functions to host services in a windowed environment, subscribe to other services, and publish its services. The framework (see Figure 1) consists of the Multi-Function Workstation, a DAF application which provides the interfaces necessary for service interaction, and web-services which are services to provide functions for computing graph metrics, process SPARQL Queries, and provide dimension reduction for understandability. The web-services are connected through the DAF to the rest of the MFWS components such as the mapping service, links visualizer, ontology viewer, and message viewer.

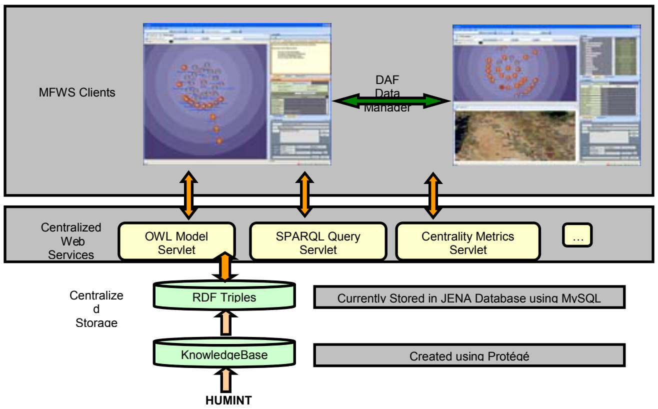

Figure 1: Technology Framework

Data Representation

RDF/OWL is a powerful method to represent data and the relationships between them. The representation schema developed is based on triples, a well known method that utilizes object-predicate-object triples to represent nodes and the links between them. Part of our research to date is in automated development of the triple store. Extraction technology is essential to this effort. Presently, we are working to automatically load extracted triples into our knowledge base and database scheme.

Our knowledge base is based on an ontology **[7].** Created from a data modeling effort, the major concepts in our ontology are individuals, organizations, events, locations, and time. Given a knowledge base, the analyst or system developer can formulate inference rules such as: If person A is related to person B and person C knows person B, then person A is associated with person C **[8].**

Knowledge bases are useful for generating graphs, since relationships are explicit. Given these explicit relationships, it is trivial to draw the resulting network, as shown in Figure 2. The use of graphs in social network analysis **[9]** provides a sound theoretical basis for understanding the relationships between entities. Graphs have properties such as degree, betweenness, and closeness centrality **[9]** which are easy to calculate for small graphs but provide important information.

Currently we build the knowledge base manually. A trained operator performs manual extraction from unstructured text, and inserts the extracted facts such as *PersonA is a member of OrganizationA* into the knowledge base. Automating this process is key to enabling automated fusion, and is an on going research project. Currently, we are using the extracted results from the Soft Target Exploitation and Fusion Army Technology Objective effort and experimenting to determine the best method to insert the facts into the knowledgebase.

Richard Antony **[10]** has developed a method for describing fusion functions which provides a framework for developing methods to solve fusion problems. He uses eight canonical forms to describe the relationships between three actors: Entity (E), locations (L), and Time (E). This taxonomy allows the categorization of problems such as {same E, near L, near T}, characteristic of a radar tracking problem, and {different E, not near L, near T}, characteristic of a cell phone call **[11].** The Antony Fusion Forms provide a consistent framework to develop algorithms based on context. We believe this method will provide a robust means to build services which are useful and deterministic.

Visualization

Presenting information to the user is an important function in any interactive system. Ease of use, flexibility, reliability, and timeliness of results are some features of a good interface.

High-dimensional data, with attributes in the hundreds, are difficult for humans to understand. Computer graphics method to pan, zoom, and rotate alleviate this some, but are insufficient. We are using a method originally developed in social sciences for exploratory data analysis, Multi-Dimensional Scaling (MDS**) [12]** to reduce the dimensionality of the vector space to a two-or-three dimensional space which is easier to understand.

Although MDS techniques have been around for a long time, they have not been applied to complex information sources such as HUMINT. We are using MDS on our datasets to determine how the technique assists in discovering HVI's. To date, we have created a notional database of individuals with different characteristics (terrorist, neural, friendly, common criminal) and analyzed it using MDS. The technique correctly clusters individuals who share some of his characteristics. This is one technique we are exploring, and will develop others as we gain experience in using them in a service oriented architecture.

We are working with Dr. James Llinas at the State University of New York, Buffalo, to research how graph matching can be used to discover HVI's **[13].** A graph

matching algorithm, called Truncated Search Tree (TruST), uses heuristic search to mitigate the NP-Hard nature of the problem **[14].** TruST finds the nearest match to a given template graph. Thus, given a candidate terrorist network from a knowledgebase query, TruST can be used to compare this network against known networks to discover a match.

User Directed Fusion

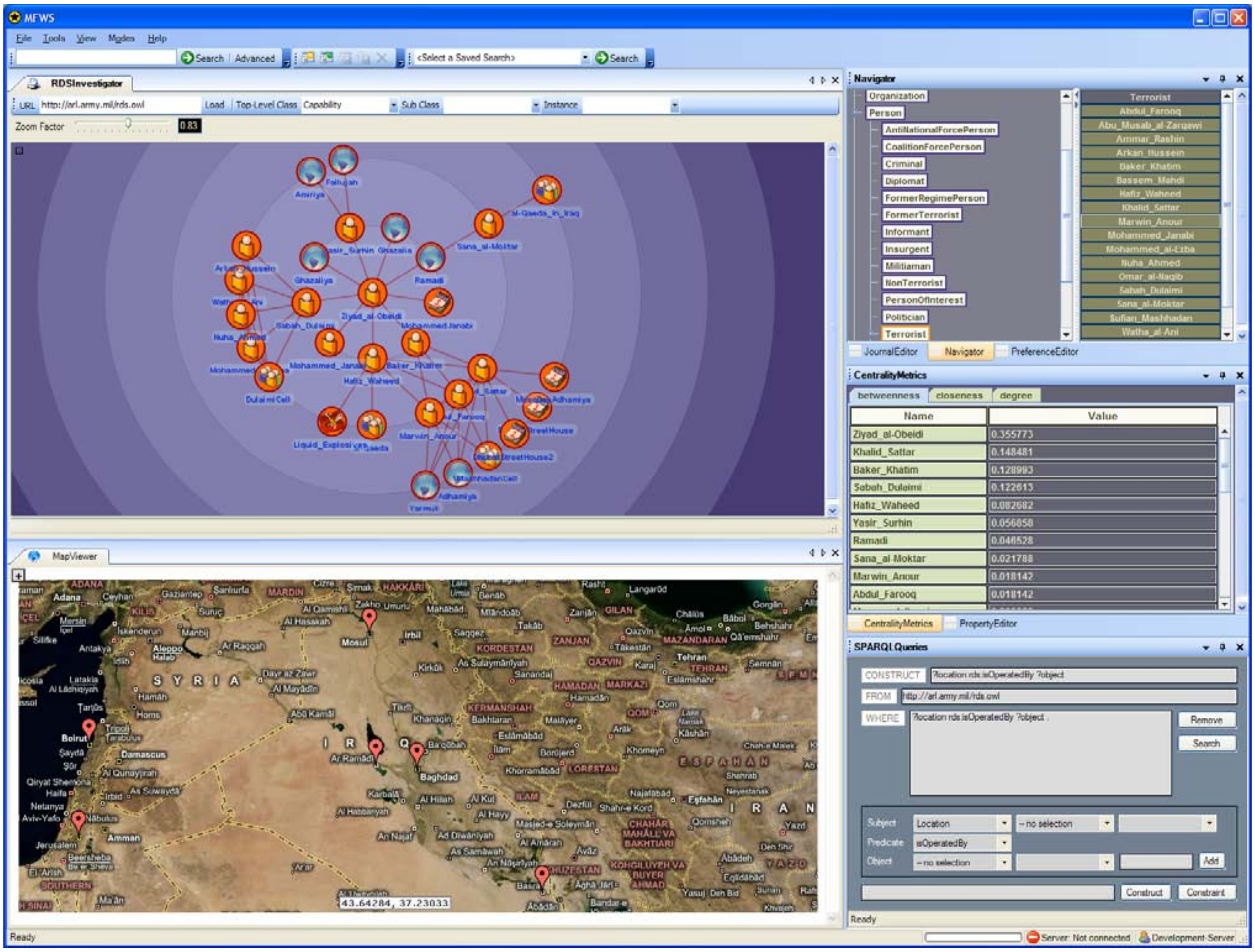

Figure 2: Prototype Analyst Workstation

The prototype system we developed (Figure 2) is a Proof-of-Concept Analyst Workstation for discovering relationships. Composed of a user-interface, database, and services, the system allows a user to make queries, visualize results using link diagrams, and view messages and ontological form. The interface addresses the needs of an analyst performing an investigation. Given a list of HVI's, the investigator may attempt to refine an earlier theory, uncover unknown HVI's, or look for links between two seemingly unrelated events. Given a suitable knowledge base, this interface allows the analyst to perform these tasks. First, a query on the HVI of interest yields (HVIi) a link diagram of other people, locations, and events with which he is associated. Computation of centrality metrics on the network reveals that another entity, probable HVI (pHVI) has a high betweenness value between two cells. A subsequent query on pHVI reveals an association with an event that was not on the network for HVIi. The analyst opens the notepad feature and notes the new link. The analyst may also enter this information into the knowledgebase explicitly. Using the MDS service, pHVI clusters not with friendlies but with terrorist. The analyst might recommend that pHVI be classified as an HVI.

Summary

We are developing a technology to provide automation support to Level 2 fusion of HUMINT. ARL is developing a framework and services, and is working

with external researchers to provide automation assistance to the analyst. This research is in its early stages, as soft fusion is a relatively new science. The metrics remain undefined at this time, so it is premature to assess its effectiveness quantitatively. However, it is clear that users of soft fusion must reach the level of confidence that hard fusion has achieved if it is to truly become a useful tool in the intelligence arsenal.

Acknowledgements

This research is supported by Soft Target Exploitation and Fusion Army Technology Objective.

References

[1] D. H. Petraeus and J. F. Amos, *FM 3-24:Counterinsurgency*, Headquarters, Department of the Army, Washington, D.C., 2006.

[2] White, Jr. F.E., Data *Fusion Lexicon, Joint Directors of Laboratories*, Technical Panel for C3, Data Fusion Sub-Panel, Navel ocean Systems Center, San Diego, CA, 1987

[3] FM 6-30

[4] Biermann, J, *Some Experience with Experimental High Level Fusion Systems*, FGAM-FKIE, 53343 Wachtberg, Germany

[5] Hecking, Matthias, Content *Analysis of HUMINT Reports*, http://www.dodccrp.org/events/2006_CCRTS/html/papers/012.pdf, accessed 10 March 2009.

[6] OWL Web Ontology Language Reference, www.w3.org/TR/owl-ref.

[7] Apenyo, Kofi, Battlespace *Relationship Discovery Ontology*, Army Research Laboratory, ARL-TR-4444, May 2008.

[**8**] Hand, David, Mannila, H., Smyth, P., *Principles of Data Mining*, The MIT Press, Cambridge, Massachusetts, 2001.

[9] Wasserman, Stanley, Faust, K., *Social Network Analysis – Models and Applications*, Cambridge University Press, 1994.

[10] Antony, R, *Principles of Data Fusion Automation*, Artech House, Inc., Boston, 1995.

[11] Antony, R. *Context & Fusion Support System (CFS2) Functional Highlights*, Briefing, January 2008.

[12] Borg, Ingwer, Groenen, Patrick J.F., *Modern Multidimensional Scaling – Theory and Applications*, Springer, 2005.

[**13**] Llinas, Jim, *Future Research Possibilities*, Briefing, Center for Multisource Fusion, March 2008.

[14] Sudit,Moises, Stotz, A., Crassidis, J. Aggamemnon, N. *Hierarchical High Level Information Fusion (H2LIFT)*, CALSPAN UB RESEARCH CENTER BUFFALO NY, Source Code: 398685 Collection: Technical Reports, AD Number: ADA486330.